2-1-2004

A Case Study of Four Location Traces

Guanling Chen
Dartmouth College

David Kotz
Dartmouth College

A Case Study of Four Location Traces

Guanling Chen and David Kotz

Dept. of Computer Science
Dartmouth College
Hanover, NH, USA 03755
{glchen, dfk}@cs.dartmouth.edu

Dartmouth Computer Science Technical Report TR2004-490

Abstract. Location is one of the most important context information that an ubiquitous-computing application may leverage. Thus understanding the location systems and how location-aware applications interact with them is critical for design and deployment of both the location systems and location-aware applications. In this paper, we analyze a set of traces collected from two small-scale one-building location system and two large-scale campus-wide location systems. Our goal is to study characteristics of these location systems ant how these factors should be taken into account by a potentially large number of location-aware applications with different needs. We make empirical measurements of several important metrics and compare the results across these location systems. We discuss the implication of these results on location-aware applications and their supporting software infrastructure, and how location systems could be improved to better serve applications' needs. In places where possible, we use location-aware applications discussed in existing literatures as illustrating examples.

1 Introduction

In recent years context-aware computing has been identified as a key enabling technology [22,6] to realize ubiquitous-computing vision [27]. Among all kinds of context information, location is the one that made the biggest impact so far. The locality of the objects and users is immediately useful to a large set of applications and based on which we also expect to infer high-level context such as meeting, with the assistance from other inputs. The recognition of the importance of location context results in numerous active research in location-provision systems [25,28,8,1,20,4,21,9,16] and location models [17,19,26,12,3,11,14].

While some evaluation of experimental location systems are available, we have not seen much reports on detailed analysis of data trace collected from a deployed location system in daily usage. Our work fills this vacancy and presents a first attempt to quantify the data quality and characterize the user movement pattern based on a 67-day location trace from an off-the-shelf locating system. While our study is inevitably subjected to the relatively small scale and biased towards certain user population, the results shall serve as a feedback on how location system behaves and a guideline for the design of location-aware applications in general.

In next tow sections we describe localization systems that can provide location information in our department building with relative fine granularity (Section 2), and on campus with coarse granularity (Section 3). We present the main results in Section 4. We discuss the related work in Section 5 and conclude in Section 6.

2 Versus

We installed a commercial locating system, VIS (Versus Information System) [1] in our department building. Our installation of VIS contains 79 IR and 7 RF (operates at 433.92 MHz) ceiling-mounted sensors, which are wired into 4 collectors (one collector handles up to 24 sensors). The collectors are then daisy-chained into a concentrator which interfaces with the Ethernet.

An IR sensor has a range of 15 feet and a RF sensor has a range of 90 feet. A personnel badge periodically emits IR and RF signals (42 bit packet with 16 bit ID), which are picked up by IR and RF sensors respectively. The packets are forwarded to collectors for re-formatting and finally relayed to the concentrator. A badge contains an IR transmitter, a RF transmitter, a motion sensor, and optionally a push button. The badge emits a RF signal about every 2 minutes, while it emits an IR signal more frequently (about every 3.5 seconds) when it is on the move than when it is stationary (about every 4 minutes). The RF channel serves as a backup indicating the badge is still in range (180 feet diameter with RF sensor as the center) even though the IR signals might be lost (for instance, due to IR's line-of-sight problem). The output of the motion sensor in the badge is used to adjust how often to send IR signals, as a way to prolong the battery's life time.

Our 3-story department building has about 60 offices, labs, and classrooms. We deployed 61 IR sensors in most rooms (except one office due to resident's request, four machine and storage rooms, and six restrooms), and 18 IR sensors to cover hallways and stairways. Several large labs were partitioned into 2 or 3 zones with one IR sensor installed for each zone. The 7 RF sensors were distributed evenly to cover the whole building.

2.1 Trace collection

We captured the output of the Versus concentrator during an academic term (January 5 to March 12, 2003). During the period for this study, both authors wear a badge "all the time". Although we distribute the badges to several other graduate students, their wearing of badges tends not consistent. There were 12 students attended a "pervasive computing" seminar during the term taught by one author. Each of them was given a badge, and they were divided into six groups to develop location-aware applications. About half of the course students have office in our building and the others do not. We also attached badges to printer, chairs, laptops, and so on. Thus the peak workload for the VIS was about 25 badges in building, and the peak load for one sensor is about 10 badges (for example, on final project demo day).

[1] http://www.versustech.com

Our 67-day trace data contains about 1.5 millions of records, each of which represents one sensor sighting of one badge. The fields of one record are shown in Table 1. There are four holes in our collected data, one on January 13 due to a software upgrade counted for 5 minutes and the other on February 25 due to a hardware upgrade counted for 2 hours. On March 6 and 7, we migrated the source collection process onto a dedicate server and lost couple of hours data on each of these two days. Unless explicitly stated, our study results do not include any data from these four days.

badge_num	the unique ID of the sighted badge
col_num	the unique ID of the collector for the reporting sensor
sen_num	the ID of the sensor under that collector
tcount	the sequence number (4 bits) of the badge signals
motion	whether the badge is in movement or stationary
button	whether the button (if the badge has one) is pressed
battery	whether the badge's battery is low
timestamp	the time when the report arrived at the logger

Table 1. List of the data fields for each record in VIS location trace.

While the RF and IR sensors are wired together in Versus system, we can think of them as two layers, one consists of IR sensors only and the other consists of RF sensors. These two layers have different characteristics such as update rate and granularity. It is interesting to study them separately and compare them shoulder-to-shoulder. Thus for our study, we separated the trace into two parts: one is generated by IR sensors, and the other by RF sensors. In the rest of paper, we refer the IR layer as **Versus/IR** and the RF layer as **Versus/RF**.

3 Campus

Our campus is compact, with over 161 buildings on 200 acres, including administrative, academic, residential, and athletic buildings. Our school installed more than 500 access points from Cisco Systems, each an Aironet model 350 [2], to provide 11 Mbps coverage to nearly the entire campus. Each access point (AP) has a range of about 130 to 350 feet indoors, so there are several APs in all but the smallest buildings. Although there was no specific effort to cover outdoor spaces, the campus is compact and the interior APs tend to cover most outdoor spaces. Thus each 802.11 device can be roughly located on campus by relating its currently associated access point.

The user population consists about 5,500 students and 1,215 full-time professors. Each year, approximately 1000 undergraduate students enter the school, and most purchase a computer through the campus computer store. Of those purchases, laptops have become increasingly dominant in recent years: 27% in 1999, 45% in 2000, and 70% in 2001. All laptops purchased in 2001 had built-in wireless support, and over 1000

[2] http://www.cisco.com/

802.11b cards have been sold over the past year to other users. In addition, all business-school students, and most engineering-school graduate students, own laptops.

We describe two approaches to infer which access point a 802.11 card is currently associated with: one is push-based (syslog) and the other is pull-based (SNMP). For both approaches, we are recording traces for the new Spring term (started on March 24, 2003). In the analysis presented in this paper, we used 24-day data sets till April 17, 2003.

3.1 Syslog trace

We configured the access points to transmit a syslog message every time a client card authenticated, associated, reassociated, disassociated, or deauthenticated with the access point (see definitions below). The syslog messages arrived via UDP at a server in our lab, which recorded all 4,316,056 of them for the 24-day period of our analysis.

Most APs contributed to the syslog trace as soon as they were configured and installed. Of the 550 APs, only 531 were represented in our trace. Although some appear never to have been used, many were misconfigured and did not send syslog messages. Since syslog uses UDP it is possible that some messages were lost or misordered. As a result of these spatial and temporal holes in the trace, some of our statistics will undercount actual activity.

Our syslog-recording server added a timestamp to each message as it arrives. Each message contained the AP name, the MAC address of the card, and the type of message:

– *Authenticated.* Before a card may use the network, it must authenticate. We ignore this message.
– *Associated.* After authentication, a card chooses one of the in-range access points and associates with that AP; all traffic to and from the card goes through that AP.
– *Reassociated.* The card monitors periodic beacons from the APs and (based on signal strength or other factors) may choose to reassociate with another AP. This feature supports roaming. Unfortunately, cards from some vendors apparently never use the Reassociate protocol, and always use Associate.
– *Roamed.* When a card reassociates with a new AP, the new AP broadcasts that fact on the Ethernet; upon receipt, the old AP emits a syslog “Roamed” message. We ignore this message; because it depends on an inter-AP protocol below the IP layer, it only occurs when a card roams to another AP within the same subnet.
– *Disassociated.* When the card no longer needs the network, it disassociates with its current AP. We found, however, that the syslog contained almost no such messages.
– *Deauthenticated.* While it is possible for the card to request deauthentication, this almost never happened in our log. Normally, the associated AP deauthenticates the card after 30 minutes of inactivity. In our log it is common to see several deauthentication messages for a widely roaming card, one message from each subnet visited in the session; we ignore all but the message from the most recent AP.

Our network does not use MAC-layer authentication in the APs, or IP-layer authentication in the DHCP server. Any card may associate with any access point, and obtain a dynamic IP address. We thus do not know the identity of users, and the IP address given to a user varies from time to time and building to building. We make the approximating assumption to equate cards with users, although some users may have multiple cards, or some cards may be shared by multiple users.

From now on, we refer to the campus-wide localization using syslog data as **Campus/syslog**.

3.2 SNMP trace

We also used the Simple Network Management Protocol (SNMP) to periodically poll the APs; 520 of the 550 APs responded to our polls. We chose to poll every 5 minutes to obtain information reasonably frequently, within the limits of the computation and bandwidth available on our two polling workstations. Each poll returned the MAC addresses of recently associated client stations, and the current value of two counters, one for inbound bytes and one for outbound bytes. Our 24-day trace period includes 35,581,599 of these SNMP records, of which we extracted 15,211,799 messages (line started with c2) for localization purpose.

In the rest of discussion, we refer to the campus-wide localization using SNMP data stream as **Campus/SNMP**.

4 Measurements and results

In this section, we present the analysis of location-update traffic. We focus the discussion on the data quality and volume (Section 4.1 and Section 4.2), load disparity across users and zones (Section 4.3), sighting intervals of users (Section 4.4), and user prevalence (Section 4.5). As a reminder, the **users** are approximated by the badges (in Versus) and 802.11 cards (in Campus). The **zones** are represented by the name of sensors (in Versus) and name of access points (in Campus).

4.1 Data quality

Unfortunately none of the four systems generates perfect location data for direct usage.

As a dedicated and commercial locating system, Versus performs most reliably and we can easily parse the data stream to obtain which badge is sighted at which sensor. Curiously, however, the system does generate some error messages reports a non-exist badge or a badge at a non-exist sensor. There are 11 such errors out of 978,469 Versus/IR trace and 57 out of 550,000 Versus/RF trace. The software part of the system has to check a human-maintained table to detect these errors.

Campus/syslog sometimes report incomplete messages (probably caused by syslog daemon), representing about 0.5 percent error rate in the trace. While it is more lightweight to transmit messages using UDP, it may cause the packets arrive out-of-order or not arrive at all. This problem is difficult to detect and impossible to recover. We ignore the warning messages about access point conditions, which represents about 23 percent of whole trace.

Campus/SNMP uses two polling stations, on each of which runs parallel polling processes. Every process polls the access points it is responsible with the interval equals

	Maximum	Minimum	Mean	Std. dev.
Versus/IR	29,341	3,093	15,531	6,993
Versus/RF	11,395	6,347	8,730	1,095
Campus/syslog	239,922	117,617	168,394	34,459
Campus/SNMP	755,531	646,927	724,371	22,884

Table 2. Statistics of daily location-update traffic.

to 5 minutes divided by the number of access points. Thus it is necessary to merge all the outputs of polling processes to have a global picture which card is associated with which access point. The results returned for a SNMP query includes statistics about AP interface cards and its associated client cards. We ignore all these and only need the list of cards (in c2 message), which represent only about 45 percent in the whole trace.

Discussion. There is a need to detect and remove errors for two Versus systems and Campus/syslog. A 23% and 55% data reduction can be gained by applying simple filtering for two Campus systems respectively. In all four systems, timestamps are put by the collection processes using machine clock when data arrives. Considering summer daytime changes, the timestamp may need to be transformed to avoid confusion (not continuous near clock changes) for timestamp-sensitive applications. Merging and correlation is necessary for reliability (Campus/syslog) and scalability (Camups/SNMP). It is possible to set up several Versus systems to cover a large building, and then the locating system has to merge all the data streams from every root concentrator to answer queries such as "Where is badge A?". *In summary, data pre-processing plays a vital role in localization systems*.

4.2 Data volume

We compare the location-update traffic generated by four systems in Table 2.

We see relatively small variation of the daily location updates for Versus/RF and Campus/SNMP. Considering the update rate is more-or-less fixed for them (pushed by badge every 2 minutes in Versus/RF and pulled by SNMP poller every 5 minutes in Campus/SNMP), this indicates that the number of daily present badges and cards has small variation too. On a typical day, Campus/SNMP produces more than 8 messages per second given total 4189 cards and 520 APs (showed up in trace).

Campus/syslog produces smaller amount of update traffic than Campus/SNMP, given the two traces are collected for same period. It is rather not surprising since syslog only produces updates when the card associates and disassociates with an access point, assuming typically this period lasts longer than 5-minute interval. It is possible, however, the card may try to associate with different AP with better signal strength or quality.

Versus/IR has a relative large variance on daily updates, due to the significant difference of update rate when the badge is in motion (every 3.5 seconds) or stationary (every 4 minutes). In theory, N badges can at least produce $(N/3.5)$ updates per second without considering missed pings caused by interference and line-of-sight problem. That is 10 updates per second with 350 active badges, and comparable to the traffic generated by Campus/SNMP. Clearly, tracking all the assets and people in a multi-site large organization will be a challenge for a location service.

Discussion. Location-aware applications reside on a mobile device will not have the resources to handle this amount of traffic. *Instead, a software infrastructure is necessary to collect, process, and disseminate the location updates to applications.* This infrastructure needs to keep up with the update arrival rate and control amount of information delivered to applications. It can shield the data pre-processing from applications, and share the results with multiple applications. There are several major research efforts specially targeted at this direction [23,6,13]. In case where data rate may still outrun the capability of infrastructure, approximation techniques have to be applied for data-stream reduction.

4.3 Load disparity

Many location-aware applications may only care about location updates from a particular zone or from a particular user, to answer queries such as “Who are in zone A?” or “Where is user X?”. We decompose the trace to show such load distribution. We compute the percentage of load updates generated by individual user, and sort them in decreasing order. We then plot the cumulative sum of the sorted vector as Y axis and cumulative sum of user number as X axis. A point is then read as “top x percent users generated y percent of location updates. We did same thing to separate the traffic for the zones.

Figure 1 shows the results for the two Versus system and Figure 2 shows the results for two Campus systems. In Versus/IR 8% sensors received about 95% updates, indicating about 7 hot zones in our building. In fact, these are the offices of the badge owners. Versus/RF, however, does not have such a deep curve because although the user population are concentrated on a small number of rooms, they are more distributed among the radio cells. About 40% badges generated 90% updates in Versus/IR, and they are more “active” ones such as the badges worn by authors and office chairs. By active, we mean the motion sensor in the badge reports badge movements. Active badges have faster update rate (every 3.5 seconds) than stationary ones (every 4 minutes). Interestingly, one badge attached to a laptop and another one to a water cup also generate large amount of updates. This indicates that the motion sensor in the badge is rather sensitive, so these two badges are also appear to be active when the user working on the desk. The number of updates by Versus/RF is insensitive to the badge state (active or not), since the update rate is fixed (every 2 minutes).

Figure 2 also shows load disparity across access points and cards. Campus/syslog has a knee near 10% cards generated 75% location updates, while the Campus/SNMP has a more smooth curve. Considering the 5-minute polling interval by Campus/SNMP, those 10% cards generate more than one “associated” or “reassociated” messages in a single interval. This is either due to the cards are indeed mobile, or a stationary card trying to “walk” through adjacent access points to find one that has better signal strength/quality. Load distribution across access points by Campus/syslog and Campus/SNMP is comparable, while the first one is more skewed than the latter.

Fig. 1. [Versus] Distribution of location updates across badges and sensors.

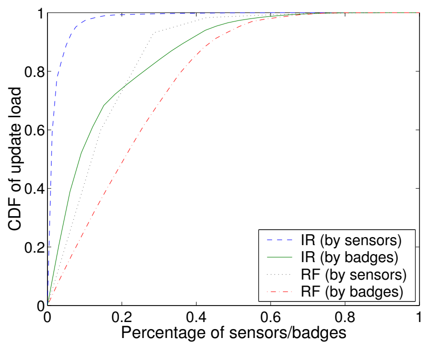
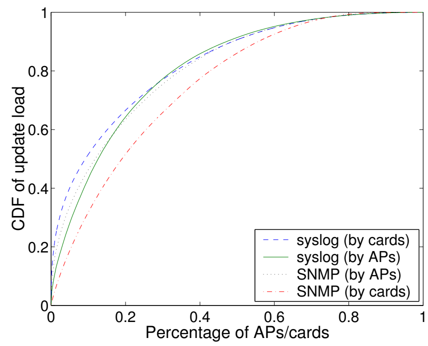


Fig. 2. [Campus] Distribution of location updates across cards and access points.



Discussion. *We observe the location-update load disparity across both zones and users in all of the four systems.* The localization systems themselves need to be aware of this situation. The sensor hierarchy, such as Versus's sensor-collector-concentrator may experience hot paths in the systems. The buffer size along these paths need to handle peak traffic not seen at the test phase. The software structure, such as a lattice-based location service [15], will also see similar phenomena. A virtual counterpart, such as an agent, for these "hot" zones and active users will experience more location updates and possibly more location queries due to popular interests. On the other hand, applications can also take advantage of such locality property. For example, popular demanded information can be cached or prefetched at a particular location.

4.4 Sighting interval

A *sighting interval* is defined as the period between two consecutive location updates for one user. For the Versus systems, we apply an cut-off period of 30 minutes. So the intervals longer than that period are discarded as they indicate the badge is out of building. We also took advantage of the motion field in the Versus events. All the intervals during the idle period (detection threshold of 3 consecutive motionless reports) are thrown away, until the badge is active again. This heuristics is used to exclude the intervals when the badge is placed on table while user is away. We apply this approach to Versus/IR and call it "Versus/IR (active)". For Campus/syslog, we use "associated" and "reassociated" to demark the boundaries of sighting. We also treat these three messages as a special location update: "deauthtication", "disassociation", and "deauthentication" with reason "inactivity". All other syslog messages are ignored in this analysis. We did not apply any cut-off period for Campus/SNMP trace. Figure 4 and Figure 3 (notice the X axis is in log scale) show the cumulative distribution of average sighting intervals (across badges and cards) for Versus and Campus systems.

The two Campus systems have a rather different sighting intervals due to the nature of their location-update mechanism. We found the mean of Campus/syslog to be

Fig. 3. [Campus] Distribution of average sighting intervals in minutes.

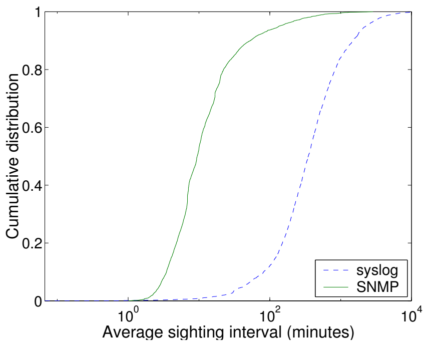


Fig. 4. [Versus] Distribution of average sighting intervals in seconds.

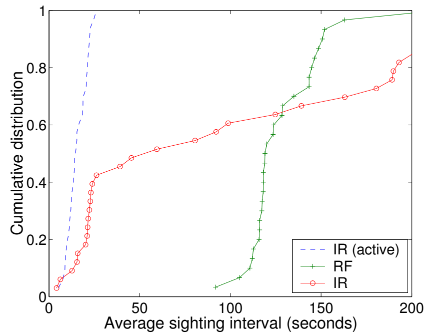


Fig. 5. [Versus] Distribution of missed pings for various badges.

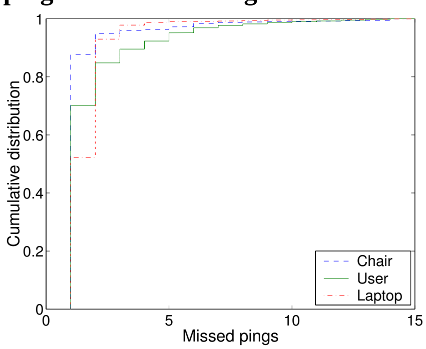
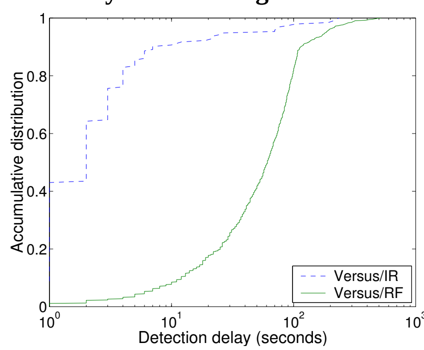


Fig. 6. [Versus] Distribution of detection delay for two badges.



383 minutes while the longest interval to be more than 157 hours. The mean of Campus/SNMP is 37 minutes, and the maximum to be about 48 hours. One possibility of interpreting the difference is that although we on purposely collected the traces over the same period time for these two systems, the set of access points shown up in the two traces are not same as we mention before. Another reason is that the SNMP 5-minute polling interval may cause it to miss some demarking association messages.

The mean of the average sighting interval for Versus/RF is 129 seconds, roughly about the fixed update rate (every 2 minutes). The mean average sighting interval for Versus/IR is 106 seconds with tail ends at 511 seconds. Notice this is much smaller than our 30-minute cut-off period, indicating that if the badge happens to leave the building, it always stays out for more than half an hour. The mean of the average sighting interval for Versus/IR (active) is 15 seconds, a magnitude smaller than previous two.

Infrared-based localization systems are well-known for its line-of-sight problems (cite Roy Want paper). The obstacles block the IR signals and potentially increase the sighting intervals. Versus system encodes a built-in sequence number in location updates that can be used to detect number of missed pings. Figure 5 shows the cumulative distribution of missed pings for three badges: one attached to office chair, one worn by a user, and another one attached to a laptop. The zero value of missed pings are not counted since they are dominant in the trace (for instance, total missed pings for the badge worn by the user is about 17 percent). We see the small number of missed pings (1 or 2) are frequent. This estimation, however, is only an approximation given the sequence number is only 4 bits. A shielded active badge will easily outrun the sequence space in $(16 * 3.5) = 56$ seconds. Starting in the March, one user starts to wear two badges, one on the wrist and the other pined on the chest. We found the chest badge lost about 18 percent pings, similar to the one on the wrist (17 percent loss).

We now do a simple analysis of detection latency. Assuming that the a random variable X, denoting the time between the badge enters a zone and the system receives its ping, uniformly distributes over 0 to n seconds. The mean of X then is $n/2$. Now assuming two statistically independent badges enter the zone simultaneously with detection time X and Y, following same uniform distribution over 0 to n seconds. It is provable that the mean of the detection-time difference of the two badges $|X - Y|$, is $n/3$. We compute the empirical detection difference of the two badges worn by the same user whenever he enters a new zone (so both badges are active), and plot the CDF in Figure 6 (note X axis is in log scale). The n for Versus/IR (active) and Versus/RF can be approximated using the mean in Figure 4 times 2, which are respectively 31 seconds and 258 seconds. The detection difference from Figure 6 are 9.6 seconds for Versus/IR and 71 seconds for Versus/RF, roughly following the $n/3$ rule. This significant detection lag should be taken into account by the applications that require spontaneous interaction using collocation information. In particular, the memory aid or active reminder like applications may miss some events and fail to take actions if two people only briefly met in the hallway.

Discussion. Sighting interval is a useful metric to measure the freshness or confidence of the location information. The more frequent the system obtains a user's location updates, the more confident of the user's current location. Active map uses image fading and frame shading during the sighting interval to degrade the information freshness gracefully [18]. The detection latency may also have impacts on many applications, such as reminder that uses collocation, the teleporting and call forwarding that follows user's current location, and so on. None of four localization systems provides an explicit location updates when the user left the covered area (the "deauthtication" and "disassociation" messages in Campus/syslog are not reliable). Thus applications have to infer this situation by applying some threshold on how long there is no update before the user's location becomes unknown. If not arbitrary, this threshold often relates to the expected value of sighting interval of that localization system.

The sighting interval is a complicated metric and jointly determined by the localization system design, the characteristics of deployed environment, and the user behavior. While in Versus/IR the percentage of missed pings is similar by either wearing the badge on the chest or the wrist, it is conceivable that a user wears the badge at the bell will suffer more lost pings. Depending on the arrangement of access points, where the card is located, and the environmental interference, the card may cause more location updates by trying to associate with an access point with best signal quality. While Cam-

Fig. 7. [Versus/IR] Scatterplot of prevalence of badges.

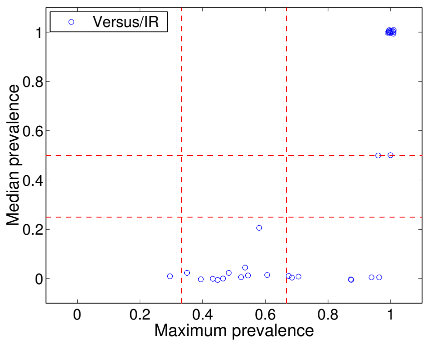
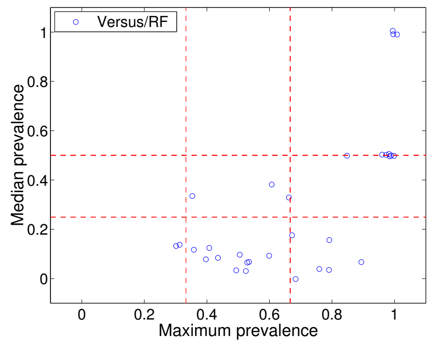


Fig. 8. [Versus/RF] Scatterplot of prevalence of badges.



pus/SNMP has a fixed 5-minute polling interval, a card can be seen much less or longer than 5 minutes if it moves around. Versus may report arbitrary long sighting intervals if the badge is taken out of building, or put in the drawer by the user, or the battery died without being noticed.

4.5 Prevalence

Prevalence of a zone in a user's trace is the measure of the fraction of time that the user spends with that zone. So for one user, there is one prevalence value for every zone she ever visited in the trace, and the sum of all the prevalence values for that user is 1. This metric is defined and used to characterize user mobility in a cooperate wireless LAN study [2]. In their study, a user is categorized as "highly mobile, somewhat mobile, regular, occasional mobile, and stationary" by segmenting maximum and median prevalence into several bins and see which bin that user belongs to. We compute the maximum and median prevalences for each user, and Figure 7 - Figure 10 show the scatterplots of every user for the four localization systems. The segmentation lines are draw in dashed lines. To prevent overplotting of same valued points, we added randomized jitter in range of $[-0.01, 0.01]$ on both X and Y axis for every point.

We first look at the two plots for Versus systems (Figure 7 and Figure 8). The total time for a user, is the time the badge is active in the system. So if the badge is left on the table unused, the period of idle time is excluded from prevalence computation. Due to the definition of "median", there will be no points with median prevalence greater than 0.5, except special case - both the median and maximum prevalence are 1 (the user only visited one place). We see there are more such stationary badges in Versus/IR than Versus/RF. At first this may seem odd since if one badge is stationary in one IR zone then it should also be stationary in a bigger RF zone. The explanation is that we have overlapping RF zones, a badge's ping received by more than one RF sensors can be reported to be any one of the RF sensor received that ping, although the exact policy when and which sensor in the group should Versus choose is unclear to us. The badges with median prevalence equals to 0.5 means that they only visited two zones,

Fig. 9. [Campus/syslog] Scatterplot of prevalence of cards.

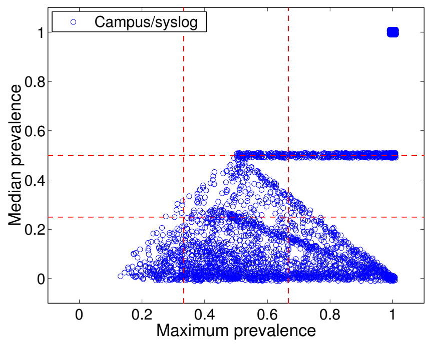
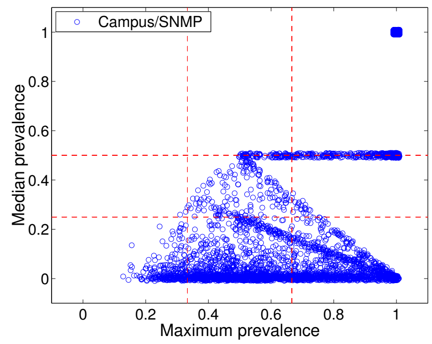


Fig. 10. [Campus/SNMP] Scatterplot of prevalence of cards.



and the maximum prevalence shows how biased the badge towards one of them. Both of authors' badges show up in the bottom right corner, indicating they spent most of time in their offices, while spent relative little time on other sensors (such as in hallways). Other badges with low prevalences indicates their time is more evenly distributed across several sensors. The badges seen by Versus/RF has relative higher prevalence, since there are only 7 radio cells.

The two plots for Campus (Figure 9 and Figure 10) bear strong similarity and show some obvious patterns. First, there are quite a few stationary cards never moved in the lifetime of the trace. Second, there are more than 10 percent of cards have 0.5 prevalence. They spent their time evenly across two access points ($P_{max} = 0.5$) or strongly tied toward one of them ($P_{max} = 1$). While some of these cards may indeed be mobile, we believe most of them just flickers between two access points when neither of their signal quality is good enough. Third, the line $P_{median} = 1 - P_{max}$ bounds all the points on the right side when the P_{median} is smaller than 0.5. This rule always holds while we skip the proof here. It is interesting to see that points actually flocks to this line. The reason we believe is that the card is flickering through 3 APs and biased strongly towards two of them, thus P_{median} approaches $1 - P_{max}$. There is another flocking line in both plots: $P_{median} = (1 - P_{max})/2$. This phenomena is caused either by the card flickering three APs while spending even time on the two besides the one with P_{max}, or the card flickering through 4 APs with one of them gets very little time.

Discussion. The above rather lengthy explanation gives the point that a card may change its associated AP while it is not actually moving and this is a significant phenomena. The network-based optimization used by the cards actually caused complication and confusion for localization systems to determine current place of a flickering card. *Using a system designed for different purpose to infer object location, sometimes conflicts our goal and complicates the task.* This suggests that our take-as-is approach to use these systems has limited usage, without advanced processing such as correlation between several APs [16]. The techniques to explicit involve user interactions during the indeterministic time have also been investigated in various applications [5].

	Maximum Prevalence (P_{max})		
Median Prevalence (P_{median})	**Low** $P_{max} \in [0, 0.33)$	**Medium** $P_{max} \in [0.33, 0.66)$	**High** $P_{max} \in [0.66, 1)$
High $P_{median} \in [0.5, 1)$	N/A	N/A (0%,0%), (2%,2%)	stationary (43%,32%), (16%,28%)
Medium $P_{median} \in [0.25, 0.5)$	N/A	regular (0%,6%), (4%,4%)	N/A (0%,3%), (1%,1%)
Low $P_{median} \in [0, 0.25)$	highly mobile (3%,6%), (6%,6%)	somewhat mobile (33%,33%), (27%,31%)	occ mobile (21%,20%), (44%,28%)

Table 3. User mobility categorization matrix of four localization systems using prevalence metric. Each entry has following format: (Versus/IR, Versus/RF),(Campus/syslog, Campus/SNMP).

We compute the percentage of the users in each mobility category (defined using median and maximum prevalence in [2]) for all four location systems and show the results in Table 3. Versus/IR and Versus/RF have the trace over the same user population over the same period, and we can see they categorize the users comparably. The main difference seems to be Versus/IR has more stationary users while Versus/RF has more highly mobile users. It is not too surprising given that a badge may be reported to show up in more than one RF cell even if it is stationary in the intersections of RF zones. We also captured the traces for Campus/syslog and Campus/SNMP over same time period with roughly same user population (the set of APs seen by two systems are slightly different, possibly some are misconfigurated and not responding to either syslog or SNMP queries). Again, these two systems divide the user population into mobility categories similarly. The obvious difference is that Campus/SNMP has less occasional mobile users but more stationary users (since the 5-minute polling interval hided some mobility). This analysis seems to allow us to conclude that *user mobility is well captured by these heterogeneous localization systems*, even they have different granularities (Versus) and location update mechanisms and rate (both Versus and Campus). The perceived user mobility between Versus and Campus also seems comparable, though they operate at different scale and track a different set of user population. One may argue that all of the four localization systems really are tracking devices, not humans. The devices are either badges, laptops, or handhelds and the users may follow certain pattern to take the device or not when they move around. However, we shall be careful to draw any conclusion here since our Versus system involves a rather small number of user population.

5 Related work

In their extensive survey [10], Hightower and Borriello define two metrics (among others) *accuracy* and *precision* to evaluate location systems. We believe our study complements their work, in the sense that we measured several other metrics that are important to location-aware applications. Our analysis is based on the traces collected from four localization systems over same period and user population, so we can compare the results among themselves. By analyzing the deployed systems with real user population, we provide feedback and guidelines for the design and deployment of both localization systems and location-aware applications.

Spreitzer and Theimer presents one of the earliest study of the innovative active badge system developed Xerox Parc [24]. The metric used by the authors is sighting interval. We obtained slightly different results due to the difference of physical characteristics of our badge systems. By taking advantage of the embedded motion sensor, we found the average sighting intervals during the badge's active time is about 15 seconds. We also measured several other metrics and discuss the impacts they may make on applications.

Balazinska and Castro recently studied a wireless LAN in a corporate site using trace obtained by SNMP [2]. They define two metrics *prevalence* and *persistence* to characterize user mobility. Using their definition and mobility categories, we computed all the prevalence values for four systems and compared the categories. Although the four location systems are quite different in the sensing granularity and location-update mechanism and rate, we found they are able to capture the user mobility in a more or less same way.

Harper reports findings about why people wear active badges based on the interviews of 44 personnels [7]. The results are more sociological than quantitative.

6 Conclusion

In this paper we analyze data traces we collected from four location tracking systems, two for an in-building deployment (IR and RF based) and two for a campus-wide deployment (push and pull based). We characterize the data quality and location-update traffic, which suggests data pre-processing is crucial for all four systems. The load disparity across the users and zones indicates both potential bottlenecks and possible optimizations using caching and prefetching techniques. The sighting intervals between consecutive updates depend on the physical characteristics of sensors and show large difference among four systems. This suggests location-aware applications must be able to cope with the interval variation and the detection latencies. By computing the prevalence metric over all four traces, we conservatively conclude that all of them capture the user mobility in a similar fashion.

References

1. P. Bahl and V.N. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. In *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies*, pages 775–784, Tel Aviv, Israel, March 2000. IEEE Computer Society Press.
2. Magdalena Balazinska and Paul Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In *Proceedings of the First International Conference on Mobile Systems, Applications, and Services*, San Francisco, CA, May 2003. USENIX Association.

3. Barry Brumitt, Brian Meyers, John Krumm, Amanda Kern, and Steven Shafer. EasyLiving: Technologies for Intelligent Environments. In *Proceedings of the Second International Symposium on Handheld and Ubiquitous Computing*, pages 12–, Bristol, UK, September 2000. Springer-Verlag.
4. Paul Castro, Patrick Chiu, Ted Kremenek, and Richard Muntz. A Probabilistic Room Location Service for Wireless Networked Environments. In *Proceedings of the Third International Conference on Ubiquitous Computing*, pages 18–, Atlanta, GA, September-October 2001. Springer-Verlag.
5. Keith Cheverst, Nigel Davies, Keith Mitchell, and Adrian Friday. Experiences of developing and deploying a context-aware tourist guide: the GUIDE project. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, pages 20–31, Boston, Massachusetts, United States, 2000. ACM Press.
6. Anind K. Dey, Gregory D. Abowd, and Daniel Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction (HCI) Journal*, 16(2-4), 2001.
7. Richard H.R. Harper. Why do people wear active badges? Technical Report EPC-1993-120, Rank Xerox Research Center, 1993.
8. Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, and Paul Webster. The anatomy of a context-aware application. In *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networking*, pages 59–68, Seattle, Washington, United States, 1999. ACM Press.
9. Mike Hazas and Andy Ward. A Novel Broadband Ultrasonic Location System. In *Proceedings of the Fourth International Conference on Ubiquitous Computing*, pages 264–, Goteborg, Sweden, September-October 2002. Springer-Verlag.
10. Jeffrey Hightower and Gaetano Borriello. Location Systems for Ubiquitous Computing. *IEEE Computer*, 34(8):57–66, August 2001.
11. Jeffrey Hightower, Barry Brumitt, and Gaetano Borriello. The Location Stack: A Layered Model for Location in Ubiquitous Computing. In *Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications*, pages 22–28, Callicoon, New York, June 2002. IEEE Computer Society Press.
12. Fritz Hohl, Uwe Kubach, Alexander Leonhardi, Kart Rothermel, and Markus Schwehm. Next century challenges: Nexus – an open global infrastructure for spatial-aware applications. In *Proceedings of the Fifth Annual International Conference on Mobile Computing and Networking*, pages 249–255, Seattle, Washington, United States, 1999. ACM Press.
13. Jason I. Hong and James A. Landay. An infrastructure approach to context-aware computing. *Human-Computer Interaction (HCI) Journal*, 16(2-4), 2001.
14. Changhao Jiang and Peter Steenkiste. A Hybrid Location Model with a Computable Location Identifier for Ubiquitous Computing . In *Proceedings of the Fourth International Conference on Ubiquitous Computing*, pages 246–, Goteborg, Sweden, September-October 2002. Springer-Verlag.
15. R. José and N. Davies. Scalable and Flexible Location-Based Services for Ubiquitous Information Access. In *Proceedings of the First International Symposium on Handheld and Ubiquitous Computing*, pages 52–, Karlsruhe, Germany, September 1999. Springer-Verlag.
16. Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Lydia E. Kavraki, Dan S. Wallach, and Guillaume Marceau. Robotics-based location sensing using wireless ethernet. In *Proceedings of the Eighth Annual International Conference on Mobile Computing and Networking*, pages 227–238, Atlanta, Georgia, USA, 2002. ACM Press.
17. Ulf Leonhardt and Jeff Magee. Multi-sensor location tracking. In *Proceedings of the Fourth Annual International Conference on Mobile Computing and Networking*, pages 203–214, Dallas, Texas, United States, 1998. ACM Press.

18. Joseph F. McCarthy and Eric S. Meidel. ActiveMap: A Visualization Tool for Location Awareness to Support Informal Interactions. In *Proceedings of the First International Symposium on Handheld and Ubiquitous Computing*, pages 158–, Karlsruhe, Germany, September 1999. Springer-Verlag.
19. Giles John Nelson. Context-aware and location systems. PhD thesis, Clare College, University of Cambridge, January 1998.
20. Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. The Cricket location-support system. In *Proceedings of the Sixth Annual International Conference on Mobile Computing and Networking*, pages 32–43, Boston, Massachusetts, United States, 2000. ACM Press.
21. Cliff Randell and Henk Muller. Low Cost Indoor Positioning System. In *Proceedings of the Third International Conference on Ubiquitous Computing*, pages 42–, Atlanta, GA, September-October 2001. Springer-Verlag.
22. B. Schilit, N. Adams, and R. Want. Context-aware computing applications. In *Proceedings of the First IEEE Workshop on Mobile Computing Systems and Applications*, pages 85–90, Santa Cruz, CA, December 1994. IEEE Computer Society Press.
23. Bill Schilit and Marvin Theimer. Disseminating active map information to mobile hosts. *IEEE Network*, 8(5):22–32, 1994.
24. Mike Spreitzer and Marvin Theimer. Providing location information in a ubiquitous computing environment (panel session). In *Proceedings of the 14th ACM Symposium on Operating System Principles*, pages 270–283, Asheville, North Carolina, United States, 1993. ACM Press.
25. Roy Want and Andy Hopper. Active badges and personal interactive computing objects. *IEEE Transactions on Consumer Electronics*, 8(1):10–20, February 1992.
26. Andrew Martin Robert Ward. Sensor-driven computing. PhD thesis, Corpus Christi College, University of Cambridge, May 1999.
27. Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, January 1991.
28. Jay Werb and Colin Lanzl. A positioning system for finding things indoors. *IEEE Spectrum*, 35(9):71–78, September 1998.